



```

RESULT 2
E311_ADE06 STANDARD; PRT: 101 AA.
AC 055653;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Early E3A 11.6 kDa glycoprotein.
OS Human adenovirus type 6.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
NCBI_TaxID=10534;
[1]
RP SEQUENCE FROM N.A.
RA Reichmann H., Schaarschmidt E., Geiseler B., Hausmann J., Ortman D.,
BA Bauer U., Flunker G., Seidel W.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL NUCLEAR MEMBRANE PROTEIN.
CC -1- PTM: N-GLYCOSYLATED AND PROBABLY ALSO O-GLYCOSYLATED (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER AD2 E3A 11.6
CC kDa, AD5 E3A 10.5 kDa, AD3 E3 9 kDa, AND AD7 E3 7.7 kDa.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y16037; CAA75991.1;
CC DR Early protein; Glycoprotein; Transmembrane.
CC KW TRANSMEM 41 62 N-LINKED (GLCNAC... ) (POTENTIAL).
CC FT CARBOHYD 14 14 N-LINKED (GLCNAC... ) (POTENTIAL).
CC SQ SEQUENCE 101 AA; 11613 MW; E219000935E3B4B CRC64;

Query Match 39.0%; Score 30; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.5e-24;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 MMFSTALMFVCLITIMLCLRRARRP 70
DB 41 MMFSTALMFVCLITIMLCLRRARRP 70

RESULT 3
E311_ADE05 STANDARD; PRT: 93 AA.
AC P17590;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Early E3A 10.5 kDa glycoprotein.
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
NCBI_TaxID=28285;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE-85092386; PubMed-2981456;
RA Cladars C., Wold W.S.M.;
RT "DNA sequence of the early E3 transcription unit of adenovirus 5.";
RL Virology 140:28-43(1985).
[2]
RN COMPLETE GENOME.
RP MEDLINE-92087470; PubMed-1727603;
RX Chroboczek J., Bieper F., Jacrot B.;
RT "The sequence of the genome of adenovirus type 5 and its comparison
RT with the genome of adenovirus type 2.";
RL Virology 186:280-285(1992).
CC -1- SUBCELLULAR LOCATION: INTEGRAL NUCLEAR MEMBRANE PROTEIN.
CC -1- PTM: N-GLYCOSYLATED AND PROBABLY ALSO O-GLYCOSYLATED.

```

```

CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER AD2 E3A 11.6
CC kDa, AD5 E3A 10.5 kDa, AD3 E3 9 kDa, AND AD7 E3 7.7 kDa.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M73260; NOT ANNOTATED_CDS.
CC DR EMBL: X03002; CAA26784.1;
CC DR PIR: A05245; ERA053.
CC KW Early protein; Glycoprotein; Transmembrane.
CC FT TRANSMEM 34 35 N-LINKED (GLCNAC... ) (POTENTIAL).
CC FT CARBOHYD 3 3 N-LINKED (GLCNAC... ) (POTENTIAL).
CC SQ SEQUENCE 93 AA; 10523 MW; 008AD087AD45A8F CRC64;

Query Match 31.2%; Score 24; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 2.8e-18;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 MMFSTALMFVCLITIMLCLRR 64
DB 34 MMFSTALMFVCLITIMLCLRR 57

RESULT 4
ATP8_ORNAN STANDARD; PRT: 69 AA.
AC Q36453;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).
GN ATP8 OR ATP8.
OS Ornithorhynchus anatinus (Duckbill platypus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
NCBI_TaxID=9258;
[1]
RP SEQUENCE FROM N.A.
RX TISSUE-Heart, and Liver; PubMed-8919867;
RX MEDLINE-97077300; PubMed-8919867;
RA Janke A., Gemmell N., Feldmaier-Fuchs G., von Haeseler A.;
RA Paabo S.;
RT "The mitochondrial genome of a monotreme -- the platypus
RT (Ornithorhynchus anatinus).";
RL J. Mol. Evol. 42:153-159(1996).
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF1) SUBUNIT OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X83427; CAA58459.1;
CC DR InterPro: IPR001421; ATP-synt_8.
CC DR InterPro: IPR003236; Mamm_mito_ATPase_8.
CC DR Pfam: PF00895; ATP-synt_8; 1.
CC DR ProDom: PD001090; Mamm_mito_ATPase_8; 1.
CC KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
CC FT TRANSMEM 8 24 POTENTIAL.
CC SQ SEQUENCE 69 AA; 8288 MW; B704FDA339643C20 CRC64;

```

Query Match 9.1%; Score 7; DB 1; Length 69;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 PPSLLQ 75  
|||||  
Db 62 PPSLLQ 68

RESULT 5  
ID Y107\_NPVAC STANDARD; PRT: 110 AA.  
AC P41660;  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hypothetical 12.5 kDa protein in H65-PK2 intergenic region.  
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
NC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=46015;  
RN [1]  
RP SEQUENCE FROM N.A.

RA STRAIN-C6; PubMed=8030224;  
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;  
RT "The complete DNA sequence of Autographa californica nuclear  
RT polyhedrosis virus."  
RT Virology 202:586-605(1994).  
RL -1- SIMILARITY: TO CORRESPONDING ORF IN OPMNPV. ORF106 AND ORF107 OF  
CC ACNMPV ARE FOUND AS A SINGLE ORF (ORF107) IN OPMNPV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: L22858; AAA66737.1; -  
CC KW Hypothetical protein.  
CC SO SEQUENCE 110 AA; 12547 MW; 27B17C9DB21204FB CRC64;

Query Match 9.1%; Score 7; DB 1; Length 110;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 ARPPSL 73  
|||||  
Db 60 ARPPSL 66

RESULT 6  
ID WRN\_MOUSE STANDARD; PRT: 1401 AA.  
AC O09053; O09050; O96242;  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Werner syndrome helicase homolog.  
GN WRN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-BAB/C; TISSUE=Testis, and Spleen;  
RX MEDLINE=97288537; PubMed=9143515;

RA Imamura O., Ichikawa K., Yamabe Y., Goto M., Sugawara M.,  
RA Furuchi Y.;  
RT "Cloning of a mouse homologue of the human Werner syndrome gene and  
RT assignment to 8A4 by fluorescence in situ hybridization."  
RL Genomics 41:298-300(1997).

RN [2]  
RP SUBCELLULAR LOCATION.  
RX MEDLINE=96284027; PubMed=9618508;  
RA Marciniak R.A., Lombard D.B., Johnson F.B., Guarente L.;  
RT "Nucleolar localization of the Werner syndrome protein in human  
RT cells."  
RL Proc. Natl. Acad. Sci. U.S.A. 95:6887-6892(1998).

RN [3]  
RP SEQUENCE FROM N.A.  
RA Paepker B.W., Gayle M., Brady W., Swartz A., Gillett L.A., Allsch R.S.,  
RA Mulligan J., Galas D., Fu Y.-H.;  
RT "Genomic structure of the human Werner's gene and cloning of its mouse  
RT homology."  
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: MAY BE INVOLVED IN THE CONTROL OF GENOMIC STABILITY.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECO SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 HRDC DOMAIN.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----

DR EMBL: D86527; BAA20270.1; -  
DR EMBL: D86526; BAA20269.1; -  
DR EMBL: AF091215; AAC78077.1; -  
DR MGD: MGI:109635; Wpn.  
DR InterPro: IPR002562; 3\_5\_exonuclease.  
DR InterPro: IPR001410; HRDC.  
DR InterPro: IPR002121; HRDC.  
DR InterPro: IPR001650; Helicase\_C.  
DR Pfam: PF01612; 3\_5\_exonuclease; 1.  
DR Pfam: PF00270; DEAD; 1.  
DR Pfam: PF00271; Helicase\_C; 1.  
DR Pfam: PF00570; HRDC; 1.  
DR SMART: SM00474; 35XOC; 1.  
DR SMART: SM00487; DEXOC; 1.  
DR SMART: SM00490; HELIC; 1.  
DR SMART: SM00341; HRDC; 1.  
KW Helicase; ATP-binding; Nuclear protein.  
FT NP\_BIND 535 542 ATP (BY SIMILARITY).  
FT SITE 632 635 DEAD BOX.  
FT FT 1115 1194 HRDC.  
FT DOMAIN 1387 1390 POLY-SER.  
FT CONFLICT 101 101 N -> S (IN REF. 3).  
FT CONFLICT 228 228 V -> A (IN REF. 3).  
FT CONFLICT 250 250 L -> S (IN REF. 3).  
FT CONFLICT 452 452 M -> V (IN REF. 3).  
FT CONFLICT 459 459 K -> T (IN REF. 3).  
FT CONFLICT 468 468 C -> R (IN REF. 3).  
FT CONFLICT 619 619 K -> Q (IN REF. 3).  
FT CONFLICT 800 800 Q -> K (IN REF. 3).  
FT CONFLICT 1021 1021 L -> S (IN REF. 3).  
FT CONFLICT 1145 1145 A -> T (IN REF. 3).  
FT CONFLICT 1181 1182 VG -> LE (IN REF. 3).  
FT CONFLICT 1252 1252 V -> A (IN REF. 3).  
FT CONFLICT 1308 1308 I -> L (IN REF. 3).  
FT CONFLICT 1356 1356 V -> A (IN REF. 3).  
SO SEQUENCE 1401 AA; 157256 MW; 949060992467FBBC CRC64;

Query Match 9.1%; Score 7; DB 1; Length 1401;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 PSLILO 75  
DB 1024 PSLILO 1030

## RESULT 7

HRK\_MOUSE STANDARD: PRT: 92 AA.  
ID HRK\_MOUSE  
AC P70678;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Activator of apoptosis harkiri (Neuronal death protein DP5).  
GN HRK OR DP5 OR BID3.  
OS Mus musculus (Mouse), and  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090, 10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Imai, K.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACTIVATES APOPTOSIS AND INTERACTS SELECTIVELY WITH  
CC SURVIVAL-PROMOTING PROTEINS BCL-2 AND BCL-XL (BY SIMILARITY).  
CC -1- DOMAIN: THE BH3 DOMAIN IS REQUIRED FOR THE INDUCTION OF CELL  
CC DEATH (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 3 (BH3).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D83697; BAA12065.1; -  
CC DR EMBL: D83698; BAA12066.1; -  
CC DR MGD: MGI:1201608; B1d3.  
CC DR InterPro: IPR000712; Bcl.2.  
CC DR PROSITE: PS01255; BH3.1.  
CC KW Apoptosis.  
CC KM DOMAIN  
CC FT SEQUENCE 92 AA: 10078 MW: 95760B9E8A8F1DE CRC64:

Query Match 7.8%; Score 6; DB 1; Length 92;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 RRRARP 69  
DB 50 RRRARP 55

## RESULT 8

VNRP\_POPMV STANDARD: PRT: 121 AA.  
ID VNRP\_POPMV  
AC 002123;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE 14 kDa protein (putative nucleic acid-binding protein).  
OS Poplar mosaic virus (isolate ATCC PV275) (PMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlaviruses.  
OX NCBI\_TaxID=31709;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE:9233381; PubMed-1629709;  
RA Henderson J., Gibbs M.J., Edwards M.L., Clarke V.A., Gardner K.A.,

RA Cooper J.I.;  
RT "Partial nucleotide sequence of poplar mosaic virus RNA confirms its  
RT classification as a carlavirus.";  
RL J. Gen. Virol. 73:1887-1890(1992).  
CC -1- FUNCTION: MAY ACT AS A REGULATORY FACTOR DURING VIRAL  
CC TRANSCRIPTION.  
CC -1- SIMILARITY: TO NUCLEIC ACID-BINDING PROTEINS FROM OTHER  
CC CARLAVIRUSES.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X65102; CA46227.1; -  
CC DR EMBL: D13364; BAA02629.1; -  
CC DR PIR: J01646; J01646.  
CC DR PIR: S23876; S23876.  
CC DR InterPro: IPR002568; Carla.C4.  
CC DR Pfam: PF01623; Carla.C4.1.  
CC FT Zinc-finger; DNA-binding.  
CC FT ZN-FINGER 62 83 C4-TYPE (POTENTIAL).  
CC FT SEQUENCE 121 AA: 14451 MW: 705205527BF7D3CF CRC64:

Query Match 7.8%; Score 6; DB 1; Length 121;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 63 RRRARP 68  
DB 52 RRRARP 57

## RESULT 9

NIFW\_FRASE STANDARD: PRT: 126 AA.  
ID NIFW\_FRASE  
AC Q92510;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Nitrogenase stabilizing/protective protein, nifw.  
GN Nifw.  
OS Frankia sp. (strain EuIK1).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Frankiaceae; Frankia.  
OX NCBI\_TaxID=47227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EuIK1;  
RA Chung-Sun A., Ji-Tee K., Won-Jin K., Won-Young Y.;  
RT "Nif-gene organization and nucleotide sequences from Frankia EuIK1  
RT strain";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: May protect the nitrogenase Fe-Mo protein from oxidative  
CC damage (by similarity).  
CC -1- SUBUNIT: Homotrimer; associates with nifD (by similarity).  
CC -1- SIMILARITY: BELONGS TO THE NIFW FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF119361; AADI7267.1; -  
CC DR Nitrogen fixation.  
KM NITROGEN  
SQ SEQUENCE 126 AA: 14287 MW: 295D8C9102FD317E CRC64:

Query Match 7.8%; Score 6; DB 1; Length 126;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 TTMATL 20  
DB 80 TTMATL 85

RESULT 10  
YRN6.CAEEL STANDARD: PRT; 136 AA.  
AC 009420:

DT 01-OCT-1994 (rel. 32, Created)  
DT 01-NOV-1995 (rel. 32, Last sequence update)  
DT 01-NOV-1995 (rel. 32, Last annotation update)  
DE Hypothetical 14.2 kDa protein R07B1.6 in chromosome X.  
GN R07B1.6.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_Taxid=6239;

RA [1]  
RA SEQUENCE FROM N.A.  
RA STRAIN-BRISTOL N2;  
RA Kershaw J.;  
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z48621; CAA8543.1;  
DR Wormpep: R07B1.6; C01632.  
DR InterPro: IPR002601; C6.  
DR Pfam: PF01681; C6.1.  
KW Hypothetical protein.  
KW SEQUENCE 136 AA; 14169 MW; D7F47A83F3F81787 CRC64;

Query Match 7.8%; Score 6; DB 1; Length 136;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 TTMATL 10  
DB 31 TTMATL 36

RESULT 11  
YB9E.YEAST STANDARD: PRT; 142 AA.  
AC P38337:

DT 01-OCT-1994 (rel. 30, Created)  
DT 01-OCT-1994 (rel. 30, Last sequence update)  
DT 15-JUL-1998 (rel. 36, Last annotation update)  
DE Hypothetical 16.4 kDa protein in POP4-SHM1 intergenic region.  
GN YBR258C OR YBR1726.  
OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetiales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_Taxid=4932;  
RA [1]  
RA SEQUENCE FROM N.A.  
RA STRAIN-S288C;  
RA MEDLINE=93220397; PubMed=8465606;  
RA Doligon F., Bileau N., Crouzet M., Aigle M.;

RT "The complete sequence of a 19,482 bp segment located on the right  
RT arm of chromosome II from *Saccharomyces cerevisiae*.";  
RL Yeast 9:189-199(1993).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X70529; CAA49922.1;  
DR EMBL: Z36127; CAA85221.1;  
DR PIR: S32960; S32960  
DR SDD: S0000462; YBR258C.  
KW Hypothetical protein.  
KW SEQUENCE 142 AA; 16429 MW; 12E73F0C3DD225D7 CRC64;

Query Match 7.8%; Score 6; DB 1; Length 142;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 GLTSAL 24  
DB 71 GLTSAL 76

RESULT 12  
YC61.MYCTU STANDARD: PRT; 149 AA.  
AC Q11057:

DT 01-OCT-1996 (rel. 34, Created)  
DT 01-OCT-1996 (rel. 34, Last sequence update)  
DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Hypothetical 16.8 kDa protein RV1261C.  
GN RV1261C OR MT1299 OR MYCY50.21.

OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_Taxid=1773;  
RA [1]  
RA SEQUENCE FROM N.A.

RA STRAIN-H37RV;  
RA MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RA "Deciphering the biology of *Mycobacterium tuberculosis* from the  
RA complete genome sequence.";  
RL Nature 393:537-544(1998).

RA [2]  
RA SEQUENCE FROM N.A.  
RA STRAIN-CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weldman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RA "Whole genome comparison of *Mycobacterium tuberculosis* clinical and  
RA laboratory strains" to the EMBL/GenBank/DBJ databases.  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO M.TUBERCULOSIS RV1558.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: 277137; CAB00892.1; -  
DR EMBL: AE007005; AAK45558.1; -  
DR TIGR: MT1289; -  
DR TubeCulList: Rv1261c; -  
KW Hypothetical protein; Complete proteome  
SQ SEQUENCE 149 AA; 16756 MW; D8E22A51B5F43CF CR664;

RESULT	14
HES2_MOUSE	
ID	HES2_MOUSE
AC	054792;
DT	15-DEC-1998 (Rel. 37, Created)
DT	15-DEC-1998 (Rel. 37, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	transcription factor HES-2 (Helix and enhancer of split 2).

```

Query Match      7.88; Score 6; DB 1; Length 149;
Best Local Similarity 100.08; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Cranata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
NCBI TaxID=10090.

```

0Y      69 PSLILL 74
Db      38 PSLILL 43

RESULT: 13
RK34-SP10L
ID      RK34-SP10L      STANDARD:      PRT:      152 AA.
AC      P82244;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      50S ribosomal protein L34, chloroplast precursor.
GN      RPL34.
OS      Spinacia oleracea (Spinach).
OC      Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC      Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
OC      Caryophyllidae: Caryophyllales: Chenopodiaceae: Spinacia.
OX      NCBI_taxid=33562;
RN      [1]
RN      SEQUENCE FROM N.A., SEQUENCE OF 92-101, AND MASS SPECTROMETRY.
RC      STRAIN=CV, ALVARO; TISSUE=Leaf;
RA      MEDLINE=20435798; PubMed=10874046;
RX      Yamauchi K., Subramanian A.R.;
RT      "The plastid ribosomal proteins: identification of all the proteins in
RT      the 50 S subunit of an organelle ribosome (chloroplast).";
RL      J. Biol. Chem. 275:28465-28482(2000).
CC      -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC      -1- SUBCELLULAR LOCATION: Chloroplast.
CC      -1- TISSUE SPECIFICITY: UBIVITIOUS.
CC      -1- MASS SPECTROMETRY: MW=6767.1; METHOD=Electrospray; RANGE=92-152.
CC      -1- SIMILARITY: BELONGS TO THE L34P FAMILY OF RIBOSOMAL PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercia
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce
CC      or send an email to license@isb-sdb.ch).
CC      -----
DR      EMBL: AF238221; AAF64157.1;
DR      InterPro: IPR000271; Ribosomal_L34.
DR      Pfam: PF00468; Ribosomal_L34.1.
DR      PROSITE: PS00784; RIBOSOMAL_L34; FALSE NEG.
KW      Ribosomal protein; Chloroplast; Transit peptide; rRNA-binding.
FT      TRANSIT      1      91      CHLOROPLAST
FT      CHAIN      92      152      50S RIBOSOMAL PROTEIN L34.
SQ      SEQUENCE      152 AA: 16095 MW: 755A930DA41ADB18 CRC64;

```

RP SEQUENCE FROM N.A.  
RA STRAIN-129/J;  
RX MEDLINE-98234545; PubMed-9570950;  
RA Nishimura M., Isaka F., Ishibashi M., Tomita K., Tsuda H.,  
RA Nakaiishi S., Kageyama R.;  
RT "Structure, chromosomal locus, and promoter of mouse Hes2 gene, a  
RT homologue of Drosophila hairy and Enhancer of split.",  
RL Genomics 49:69-75(1998).  
CC -I- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH  
CC PROTEIN FOR THEIR TRANSCRIPTION.  
CC -I- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX  
CC WITH A CO-REPRESSOR PROTEIN (GRUOCH).  
CC -I- SUBCELLULAR LOCATION: Nucleat.  
CC -I- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A  
CC HELIX-INTERERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),  
CC RATHER THAN THE CANONICAL E-BOX (CANNGG).  
CC -I- DOMAIN: THE CARBOXYL-TERMINAL WRW MOTIF IS A TRANSCRIPTIONAL  
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GRUOCH, A  
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY  
CC HAIRY-RELATED PROTEINS.  
CC -I- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and/or commercial  
CC entities requires a license agreement (See <http://www.isdb-sib.ch/announce>,  
CC or send an email to [license@isdb-sib.ch](mailto:license@isdb-sib.ch)).  
-----  
DR EMDL; AB009967; BAA24091.1; .  
DR MGD; MG1:1098624; He92.  
DR InterPro; IPR003015; HLH\_Myc.  
DR InterPro; IPR001092; HLH\_dlm.  
DR InterPro; IPR003650; Orange.  
DR Pfam; PF00010; HLH; 1.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00511; ORANGE; 1.  
KW PROSITE; PS00036; HELIX\_LOOP\_HELIX; 1.  
FT Transcription Regulation; DNA-binding; Nuclear protein; Repressor.  
FT DNABIND 13 26  
FT DOMAIN 139 148  
FT DOMAIN 154 157  
FT WRW MOTIF (REQUIRED FOR ACTIVITY)  
FT (BY SIMILARITY).  
SQ SEQUENCE 157 AA; 17231 MW; 570AC67E4992EA7 CRC64;

Query Match	7.8%	Score 6;	DB 1;	Length 152;
Best Local Similarity	100.0%	Pred. No. 35;		
Matches	6;	Conservative	0;	Mismatches
			Indels	Gaps
Oy	62	LKRRRA	67	

Query Match	7.88;	Score 6;	DB 1;	Length 157;
Best Local Similarity	100.08;	Pred. No. 36;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	63	KRRRAR 68		

DB 23 KRRRAR 28

DB 23 KRRRAR 28

Search completed: June 21, 2002, 08:25:26  
Job time: 311 sec

```

RESULT 15
HES2_RAT STANDARD: PRT: 157 AA.
AC P35429;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Transcription factor HES-2 (Hairy and enhancer of split 2).
GN HES2 OR HES-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=9335886; PubMed=8354270;
RA Ishibashi M., Sasai Y., Nakanishi S., Kagayama R.;
RT Molecular characterization of HES-2, a mammalian helix-loop-helix
RT factor structurally related to Drosophila hairy and Enhancer of
RT split.
RT Eur. J. Biochem. 215:645-652(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
CC PROTEIN FOR THEIR TRANSCRIPTION.
CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
CC HELIX-INTERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CAGNAG),
CC RATHER THAN THE CANONICAL E-BOX (CANNTG).
CC -1- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO, A
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
CC HAIRY-RELATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D14029; BAA03118.1; -
CC PIR: S35037; S35037.
CC TRANSFAC: T01650; -
CC InterPro: IPR003015; HLH_Myc.
CC InterPro: IPR001092; HLH_dlm.
CC InterPro: IPR001650; Orange.
CC Pfam: PF00010; HLH.1.
CC SMART: SM00353; HLH.1.
CC SMART: SM00511; ORANGE.1.
CC PROSITE: PS00038; HELIX_LOOP_HELIX.1.
CC Transcription regulation: DNA-binding; Nuclear protein; Repressor.
CC DNA_BIND 13 26 BASIC DOMAIN.
CC FT DOMAIN 27 71 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC FT DOMAIN 139 146 PRO-RICH.
CC FT DOMAIN 154 157 WRPW MOTIF (REQUIRED FOR ACTIVITY)
CC (BY SIMILARITY).
CC SO SEQUENCE 157 AA; 17028 MW; B5D621E814AE0369 CRC64;

```

## Query Match

Best Local Similarity 7.8%; Score 6; DB 1; Length 157;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 KRRRAR 68

|||||

